DOCUMENT RESUME

ED 269 451 TM 860 298

AUTHOR Noland, Theresa Koontz; Taylor, Bob L.
TITLE The Effects of Ability Grouping: A Meta-Analysis of Research Findings.
PUB DATE Apr 86
NOTE 68p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Ability Grouping; Academic Ability; *Academic Achievement; Affective Measures; Cognitive Tests; Effect Size; Elementary Secondary Education; Instructional Program Divisions; *Meta Analysis; Racial Factors; *Scores; *Self Concept; Sex Differences; Socioeconomic Status; Statistical Data; Tables (Data); Time on Task

ABSTRACT

The study reported in this paper quantitatively integrated the recent research findings on ability grouping in order to generalize about these effects on student achievement and student self-concept. Meta-analysis was used to statistically integrate the empirical data. The relationships among various experimental variables including grade level, sex, race/ethnicity, ability level of students, the subject matter being taught, the length of time students were grouped, and the effects of ability grouping were investigated. The relationships among various study characteristics including publication data, school size data, school socio-economic status, school geographic location, school curriculum information, the number of ability levels, measurement information, experimenter information, internal validity and the effects of ability grouping were investigated. The major findings of this study, based on 720 measurements derived from experimental data in 50 studies reported between 1967 and 1983, were that students who were ability grouped had the same cognitive outcome scores as students who were not ability grouped, and had lower affective outcome scores than students who were not ability grouped. The findings indicated not only that the practice of ability grouping does not increase student achievement, but also that ability grouping has adverse effects on students' self-concept. A bibliography containing nine citations from 1971 to 1985 is included. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED269451

THE EFFECTS OF ABILITY GROUPING: A META-ANALYSIS OF RESEARCH FINDINGS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. K. Noland

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Theresa Koontz Noland
Bob L. Taylor, Professor of Education

University of Colorado
Boulder, Colorado
1986

Presented at

American Educational Research Association

Annual Conference

San Francisco, California

April, 1986

TM 860 298

ABSTRACT

"The Effects of Ability Grouping: A Meta-Analysis of Research Findings"

Noland, Theresa Koontz and Bob L. Taylor

Paper presented at American Educational Research Association Annual Conference, San Francisco, California, April 1986

In thousands of classrooms today, students are grouped for instruction according to someone's judgment of their ability. The purpose of this study was to quantitatively integrate the recent research findings on ability grouping in order to make clear generalizations about the effects of ability grouping on student achievement and student self-concept.

The methodology used to statistically integrate this body of empirical research was meta-analysis.

The relationships among various experimental variables including grade level, sex, race/ethnicity, ability level of students, the subject matter being taught, the length of time students were grouped and the effects of ability grouping were investigated.

The relationships among various study characteristics including publication data, school size data, school socio-economic status, school geographic location, school curriculum information, the number of ability levels, measurement information, experimenter information, internal validity and the effects of ability grouping were investigated.

The major findings of this study, based on 720 measurements derived from the experimental data presented in 50 studies reported between 1967 and 1983, were that students who were ability grouped had the same cognitive outcome scores as students who were not ability grouped, and had lower affective outcome scores than students who were not ability brouped.

In addition, researchers who were identified as favoring ability grouping found that ability grouping increased achievement and researchers who were identified as opposing ability grouping found the opposite. Both groups of researchers found that ability grouping had adverse effects on students' self-concept.

The findings of this study indicate that the practice of ability grouping does not increase student achievement and does damage student self-concept. School personnel should use ability grouping very judiciously if at all.

THE EFFECTS OF ABILITY GROUPING; A META-ANALYSIS OF RESEARCH FINDINGS

In thousands of classrooms today students are grouped for instruction according to someone's judgment of their ability. The educational value of this assignment to ability groups for instruction has yet to be verified. The purpose of this study was to quantitatively integrate the recent research findings on ability grouping in order to make clear generalizations about the effects of ability grouping on student achievement and student self-concept.[1]

Ability grouping is the practice of organizing classroom groups of students who have most nearly the same standing on measures or judgments of learning achievement or capability, usually on a semester or yearly basis. It is "the major mode of horizontal differentiation within schools."[2] "While most schools practice some form of ability grouping, it is particularly widespread in large, racially and economically diverse school systems and in schools serving primarily lower class or minority students."[3]

In one of the earliest reviews of ability grouping research, Miller and Otto summarized the results of twenty studies on ability grouping. As continues to be the case with recent reviewers of this large body of research, they were critical of the experimental methodology used by some researchers. They concluded that "the evidence is contradictory. . . . So far as achievement is concerned, there is no clearcut evidence that homogeneous grouping is either advantageous or disadvantageous."[4]

Fifty years later, in one of the most recent reviews, Hey concluded that "ability grouping, as now practiced, has not provided the benefits commonly foreseen by educators and may in fact be causing serious psychological and educational harm to large numbers of public school students."[5]

The reviewer of the research on ability grouping finds remarkable consistency in the inconsistency of that research. Although some individual studies continue to find evidence that ability grouping either improves student achievement or enhances student self-concept, eleven of the thirteen major reviews of the ability grouping research stated that there was no clear-cut evidence that there is an achievement advantage to be gained from ability grouping. In addition, two reviews, including the most recent review, concluded that ability grouping has a deleterious effect on the achievement of students who are ability grouped. Although the review conclusions are consistent some individual studies still show advantages for ability grouping while others do not.

Even a cursory glance at the dozens of studies of ability grouping reported since 1967 shows that a wide variety of techniques, measurement instruments, research designs, types of persons, and statistical techniques to interpret the data have been used. This large body of studies yielding contradictory or inconsistent empirical findings has allowed school policy makers to "pick and choose" among studies to support almost any position a school or school district staff wished to take on the issue of ability grouping. The conclusions drawn from a meta-analysis of the ability grouping might have the power to move social science research and schools in the direction of orderly progress and the development of a more

consistent set of answers to the fundamental questions about how schools should be organized and how students learn.

Problem Statement

The research questions asked in this study were:

1. What relationships exist between ability grouping and student achievement?

2. What relationships exist between ability grouping and student self-concept?

3. What interactions can be observed between study characteristics, experimental variables, and the effects of ability grouping?

Methodology

The meta-analysis technique used in this study was essentially a three-step process: first, an exhaustive literature search was conducted; second, the studies collected were described, classified and variables defined as fully as possible; and finally a variety of statistical, data analysis techniques were performed to examine the data as fully as possible. In a very real sense, meta-analysis is conducted as a standard research study--problem statement, literature search, data-collection, data analysis, and report of results--using individual studies (rather than individuals) as the datum points. "The written work of others (whether or not formally published) constitutes the 'raw' data for the integration."[6]

Selection of Studies

The literature search for this meta-analysis began with several studies of interest, already known, being located in the various indexes. The descriptors used to describe the studies were noted. Additional studies and their descriptors were located in this fashion. Using these descriptors, articles in journals, theses, reports to government agencies, papers from scholarly meetings, reports to foundations, and the like were located. Secondary sources were searched to locate additional primary sources. Review periodicals, various abstract and citation archives were searched, and various computerized search services were utilized. Additionally, the bibliographies of the studies located provided additional studies. From the initial search pool of 1400 articles and their bibliographies, a core of experimental studies was uncovered. Eventually, no new relevant references were found.

Obviously, 1400 studies were not used in this meta-analysis. The studies used were limited to those which fell into the following categories:

--Only studies published after the important Hobson v. Hansen (1967) case were used.

--Only studies which were experimental or quasi-experimental and which reported data on both homogeneously and heterogeneously grouped students were used.

--Only studies conducted on students in grades kindergarten through 12 were used.

--Only studies conducted in the United States were used.

Methodology

When the body of available studies had been located, the quantitative description of the characteristics and findings was undertaken. The experimental variables which were quantified, coded and recorded for each study are shown in Table 1. These variables were chosen because researchers or reviewers interested in the effects of ability grouping had at one time or another found them to be related to the outcomes of ability grouping. (All tables referred to in this paper appear at the end of the paper.)

A strength of meta-analysis is that it can show the relationship of various study characteristics to effect sizes. The study characteristics which were quantified, coded and recorded for later analysis are shown in Table 2. These study characteristics were chosen because researchers or reviewers interested in the effects of ability grouping or of experimental studies in general had at one time or another found them to be related to the outcomes of experimental studies. Several characteristics were chosen because they were of interest to this researcher.

In addition to recording study characteristics and experimental variables, the outcomes of each study were recorded. A variety of methods for computing effect size, that is, the outcomes of a study, have been proposed and used in meta-analytic studies. The most powerful method is the effect size method because "it permits the application of a broad range of techniques, from frequency distributions to multiple regressions" in the analysis of the data from each study.[7]

For these reasons, and because of its conceptual simplicity, Glass's Δ was used in this study. The basic idea of Δ is to show the

actual magnitude of the difference between the homogeneous and the heterogeneous groups in different studies on a common scale, that common metric is Δ. The most straightforward formula for obtaining Δ is

$$ES \text{ or } \Delta = (\bar{X}_E - \bar{X}_C)/S_x \quad (1)$$

where $\bar{X}_E$ is the mean score of the experimental group, and

$\bar{X}_C$ is the mean score of the control group, and

S_x is the control group standard deviation.[8]

The effect size, Δ, represents how much better or worse the homogeneous group did than the control group did on the outcome measure used by the original researcher. An effect size of Δ = +1 would mean that the homogeneous group scored one standard deviation higher than the heterogeneous group on the outcome measure used by the original researcher.

The independent variables measured in this study were the variety of outcome measures researchers had used in their experiments of the efficacy of ability grouping. Those outcomes fell roughly into two categories: cognitive and affective.

The major hypotheses tested in this study were:

1. There is no difference in the overall outcome scores of students when they are ability grouped for instruction and the outcome scores of students who are not ability grouped.

2. There is no difference in the overall cognitive outcome scores of students when they are ability grouped for instruction and the cognitive outcome scores of students who are not ability grouped.

3. There is no difference in the overall affective outcome scores of students when they are ability grouped for instruction and the affective outcome scores of students who are not ability grouped.

These hypotheses were tested even though using inferential statistical methodology and meta-analysis is still somewhat controversial. Glass posited that "[i]nferential statistics seem to work well in two instances: randomized experiments and well-designed surveys with explicit sampling techniques."[9] This meta-analysis does not meet those requirements for valid statistical inference, and the various inferential statistics, e.g., standard errors of the more important averages and 95 percent Confidence intervals, are reported in this study with the reminder to the reader that these data are heavily interdependent.

The statistical dependence of this data set--fifty studies yielding 720 effect sizes--is complex. For example, one study yielded sixty-three effect sizes (three experimental groups, several subject matter and self-concept measures and a series of testing times), while other studies yielded only one or two effect sizes. The data regarding effect sizes per study are shown in Table 3 Despite the fact that these data are heavily interdependent much important information would have been lost if the findings in each study had been averaged and then study level effect sizes used for the meta-analysis, because many, in fact most, studies used a variety of types of students, grade levels, subject matters, and investigated both cognitive and affective outcomes. For example, Table 4 shows that the effect size was 0.00 (n = 50) when findings were averaged at the study level. The range was two standard deviations narrower than when

individual effect sizes were used for analysis. Clearly, a great deal of interesting data would have been lost if study level findings had been used.

The issue of the best way to deal with this non-independence is a continuing one. No fully satisfactory solution has been proposed, and so the user of the meta-analytic data must keep in mind the non-independence problem. For "[d]espite the limitations of meta-analysis, it is currently the best available method for cumulating and integrating the results of research."[10]

Major Findings

The first part of this section presents the overall results of this meta-analysis of the ability grouping research. The second section presents the relationship of various experimental variables to the findings. The third section presents the relationship of various study characteristics to the findings of the ability grouping research.

Overall Findings

As shown in Table 5, the mean effect size for all studies comparing students who were placed in homogeneously grouped classrooms versus similar students in heterogeneous classes on all outcome measures was -0.08. The range of effect sizes was +2.57 to -2.31. The 95 percent confidence interval was -.13 to -.03. Thus the hypothesis that there is no difference in the outcome scores of students who were ability grouped for instruction and similar students who were not ability grouped can be rejected. The overall outcomes of ability grouping are negative. In other words, the

student in a heterogeneous class scored higher overall than the similar student in an ability grouped classroom.

Cognitive Findings

As shown in Table 6, the mean effect size for all studies comparing students who were placed in homogeneously grouped classes versus similar students in heterogeneous classrooms on all cognitive measures was .01. The range of effect sizes was from +2.57 to -2.31. The 95 percent confidence interval was -.07 to .10. Thus the hypothesis that there is no difference in the cognitive outcome scores of students when they are ability grouped for instruction and the scores of similar students who were not grouped cannot be rejected.

Affective Findings

As shown in Table 7, the mean effect size for all studies comparing students who were homogeneously grouped versus similar students in heterogeneous classes on all affective measures was -0.15. The range of effect sizes was 1.99 to -1.57. The 95 percent confidence interval was -0.21 to -.10. Thus, the hypothesis that there is no difference in the affective outcome scores of students when they were ability grouped and the affective outcome scores of similar students who were not ability grouped can be rejected. The affective outcomes of ability grouping were negative. Self-concept, self-esteem and other affective outcome scores were lower for those students who were ability grouped than for those students who were not.

Additional Findings

Table 8 shows the average effect sizes obtained for every type of outcome measured in the fifty studies in this meta-analysis. Cognitive outcomes are listed in the first section, followed by affective outcomes and other types of outcomes. Table 8 presents a variety of interesting data, only a sample of which are discussed here:

--Content Area Knowledge was the cognitive outcome most often tested, $\Delta = -0.04$ ($n = 98$)

--Academic Self-Concept was the most often tested affective outcome, $\Delta = -0.30$ ($n = 133$).

--Content Area Skill scores were almost a quarter of a standard deviation higher, $\Delta = .24$ ($n = 18$), for those students who were ability grouped.

--Student Attitude Toward Subject Matter Being Taught scores were almost a quarter of a standard deviation higher, $\Delta = 0.22$ ($n = 15$), for those students who were ability grouped.

--Academic Self-Concept scores were almost a third of a standard deviation lower, $\Delta = -0.30$ ($n = 133$), for those students who were ability grouped.

--Self-Esteem scores were almost four-tenths of a standard deviation lower, $\Delta = -0.36$ ($n = 67$), for those students who were ability grouped.

Experimental Variables Findings

This section presents a statistical description of how various experimental variables were related to the impact of ability grouping.

Rather than seeking statistical significance for these findings, the reader should note relative differences.

Some of the findings shown in this set of tables are discussed here:

--As shown in Table 9, none of the overall grade level average effect sizes were positive except for the Middle School level, and that finding was based on an n of only 24.

--As shown in Table 10, for grades K-3, the effect of ability grouping was essentially 0 (n = 84); for grades 4-6 the effect was -0.20 (n = 58), with the students who were not ability grouped scoring two-tenths of a standard deviation higher than the students who were ability grouped; for the Middle Level students, both Middle School and Junior High School students scored higher, +0.08 (n = 24) and +0.22 (n = 87) when ability grouped; for High School students the reverse was true, with high school students scoring more than a third of a standard deviation lower, -0.34 (n = 19), when ability grouped.

--As shown in Table 11, when the feelings of students were measured, students in all grade levels scored lower when ability grouped.

The impact of ability grouping was different for males and females. Overall findings, which indicate no difference in the impact of ability grouping on males and females, mask some important differences. As shown in Table 12, for both females and males the impact of ability grouping was negative for achievement, however, for females very few (n = 2) effect sizes were reported, while a great many cognitive effect sizes (n = 194) were reported

for groups which were 100 percent male. As shown in Table 13, when affective outcomes were measured, the negative impact of ability grouping was twice as great on males, -0.22 (n = 166), as on females, -0.11 (n = 69). Overall, the percentage of female scores in the ability group scores correlated -.06 (p = .1) with effect size.

Four hundred and sixty-nine of the effect sizes in this meta-analysis were derived from studies which included no minority students. Only thirty came from studies which had 90 percent or more minority students. Most researchers controlled for the potentially different impact of ability grouping on students from various minority groups by having almost equal numbers of minority students or no minority students in the experimental or control groups. As shown in Table 14, when the percentage of minority students in an ability grouped classroom was high, the cognitive outcome scores were positive, +0.31 (n = 18). It should be noted, however, that this finding was based on a relatively low number of effect sizes. As shown in Table 15, when the percentage of minority students in an ability grouped classroom was high, the affective outcome scores were also positive, = 0.24 (n = 15). Again, this finding was based on a low number of effect sizes.

Although four hundred and sixty-nine, 65 percent, of the effect sizes used in this meta-analysis were from studies which had no minority students in either the experimental or the control group, several studies did report samples with identified minority groups. However, the samples of

minority populations were too small to make reliable generalizations about the imapct of ability grouping on any specific minority group.

Some of the studies in this meta-analysis reported scores for students in high, average and low ability groups, while others reported only average scores for all students who were grouped versus students who were not ability grouped. As shown in Table 16, when only cognitive measures are considered, the effects of ability grouping are positive for both high and low ability students, +0.16 (n = 82) and +0.18 (n = 72) respectively, but quite negative, -0.45 (n = 36) for average ability students. In other words, average ability students scored almost one-half of a standard deviation lower on achievement-type measures when ability grouped than similar students who were not so grouped.

As shown in Table 17, when self-esteem and other affective outcomes were measured, the students in the lowest ability group suffered the greatest damage. The lowest ability level students scored more than three-tenths of a standard deviation lower, -0.35 (n = 106) when ability grouped than similar students who were not ability grouped.

Where data were available on how school personnel placed students into already existing ability groups, they were collected for analysis: 51 percent used IQ tests in the decision-making process, 50 percent used achievement scores, 43 percent used teacher recommendations, 22 percent used reading test scores, 17 percent used grade point averages, 3 percent used parent recommendations, and 24 percent used other criteria such as administrative recommendations. Most schools used a combination of two or

more criteria. The bases used for the ability grouping decisions were not related to the outcomes of ability grouping.

Perhaps some subjects are more effectively taught and student self concept is enhanced when students are ability grouped in that subject area--perhaps others are not. Almost half of the effect sizes in this meta-analysis were conducted using all the subjects in the curriculum; however, for those studies where the subject matter for which students were ability grouped was reported, the data are presented in Tables 18 and 19. On both of these tables, the category "Other" includes such subjects as Physical Education, Foreign Language and Art. For both cognitive and affective measures, when the subject was Reading, ability grouping had positive effects, +0.06 (n = 83) and +0.16 (n = 6). For all other subjects which were investigated the affective scores of students who were ability grouped for instruction were lower.

Tables 20 and 21 present the data on average effect sizes broken down by grade level, subject matter and ability level of the students. When the data are presented in this manner, most of the ability level cells have either no effect sizes or a very small number of effect sizes. Inspection of these two tables reveals that no pattern of effect sizes emerges, and no clear generalizations can be made.

The amount of time per day which the student spent in the ability group was considered likely by some researchers to have an impact on the results of ability grouping. As shown in Table 22, on measures of achievement, spending approximately one period per day in the ability grouped classroom had a negative impact, -0.27 (n = 53) on the achievement of the

ability grouped students, while spending the entire school day in an ability grouped classroom had a positive effect, +0.22 (n = 154). As shown in Table 23, on affective measures, spending 50 percent of the school day or less in an ability grouped classroom had a more negative impact, -0.31 (n = 27), than did spending 100 percent of the school day in an ability grouped classroom, -0.13 (n = 256).

Study Characteristics

A strength of meta-analysis is that it can investigate the relationships of various study characteristics to study findings. The study characteristics findings in this section are presented in the following order: publication findings, school findings, curriculum findings, measurement findings, experimenter findings, and overall study findings.

Publication Findings

Table 14 shows both how widely interest in ability grouping as a research topic has varied and how the findings of that research on ability grouping have varied over time. As the practice of ability grouping came under attack in the late 1960s and then gained respectability again in the late 1970s, the number of studies reported changed, the range of effect sizes found changed, and the effect sizes themselves also changed. During the first three years covered by this study, years of social ferment and social change in this country, there were an average of 97 effect sizes per year reported; while in the most recent three years, there were only four effect sizes per year reported—and all of these were found in a 1983 study. No ability grouping studies suitable for this meta-analysis were

found for the years 1981 and 1982. The largest average positive effect sizes, +0.38, +.51, and +0.27, were all found in years when very few investigations were reported, n = 3, 7 and 4 respectively. The largest average negative effect size, -0.28, represents a considerable number of investigations, and in fact, it occurs in the year that had the most active reporting of ability grouping studies, n = 169. The overall average effect size correlated negatively, -0.18 (p = .001) with the year the study was reported.

Studies of ability grouping and its effects were found in journals, theses and various unpublished sources such as ERIC documents. As shown in Table 25, the largest average effect size was found in studies published in journals and the smallest average effect size was found in various unpublished forms, -0.25 (n = 108) and .06 (n = 237) respectively. This finding is consistent with Glass's analysis of eleven meta-analyses which found direct evidence of a publication bias in favor of larger effect sizes.[11] Clearly, an integrated review based solely on published work would arrive at a much different conclusion than one based on an assiduous search using all forms of research, unpublished as well as published.

School Findings

Some researchers believe that various types of educational treatments such as ability grouping have different effects in different sizes of schools. Assuming that a researcher would indicate if the school(s) being used for research were other than "average" size, 69 percent of the effect sizes in this study came from schools of average size, that is, from schools which were not defined as either small or large by the researcher. The

findings do not indicate any school size by ability grouping effect.

To a certain extent schools reflect the social climate of the community in which they are located, and community wealth and education may have an impact on the effects of ability grouping on students. Again assuming that a researcher would clarify the socioeconomic status of the school if the status were definable, it was assumed that 52 percent of the effect sizes came from studies conducted in schools which were heterogeneous in socioeconomic status. The effect sizes for studies which reported the socioeconomic status of the school (or from which the socioeconomic status of the school could be inferred) and which used cognitive measures are shown in Table 26. For achievement measures, all effect sizes were positive except for schools reporting a heterogeneity in terms of social class, where ability grouping had a negative effect, -0.20 (n = 40). It is precisely in those schools where students come from a variety of social classes that a student's socioeconomic background might lead to special class placement. As shown in Table 27, regardless of the socioeconomic status of the school community, the affective impact of ability grouping was negative. Again, however, the largest negative impact was felt in schools which displayed a heterogeneity of social class backgrounds.

Geographic location of a school has also been thought to be related to the impact of various educational practices. The overall average effect sizes for various geographic locations is shown in Table 28. The inner city school location reported a positive overall impact for ability grouping, while the small town, suburban and mixed locations all showed negative overall impacts for ability grouping.

Curriculum Findings

In some of the studies in this meta-analysis, curriculum changes were made for one or more of the groups of interest. Because information about curricular changes was very sketchy at best, this variable produced no useful information.

Measurement Findings

Sixty-nine percent of the studies in this meta-analysis were from studies which used nationally published tests for measurement information. The overall effect sizes by type of measurement instrument used by the original researcher are shown in Table 29. Nationally published tests had the largest range of effect sizes, +2.57 to -2.31 (n = 499).

Some ways of measuring results are more objective than others. Glass described this as reactivity, and said the "[h]ighly reactive instruments are those which reveal or closely parallel the obvious goals of the experimenter. . . . Relatively nonreactive measures are not so easily influenced. . . ."[12] Using this definition, a scale was developed with low and non-reactive measures such as large-scale achievement testing at one end and highly reactive measures such as a teacher's judgments about a student's liking ability grouping at the high end. The effect sizes for studies using instruments of various reactivity are shown in Tables 30, 31, and 32. Not surprisingly, the largest number of affective measures were rated highly reactive, while most cognitive measures were rated as exhibiting low reactivity. However, somewhat surprisingly, the effect sizes for instruments of varying reactivity varied very little. In this study, the

reactivity of the measurement instrument had virtually no effect on study findings.

The results of an analysis of placement methods, that is, the method the experimenter used to place students into the experimental group or the control group used by the original experimenter, are shown in Tables 33 and 34. Seventy-one percent (n = 493) of the effect sizes in this meta-analysis were from studies using intact groups. For cognitive measures, students in randomly formed ability groups achieved worse than intact groups, -0.20 (n = 70) and -0.01 (n = 132) respectively. For affective measures, the students in the experimental groups which had been formed randomly had a negative effect size of only -0.04 (n = 48) while those in experimental groups which were already in existence for school purposes had a negative effect size of -0.18 (n = 325).

Experimenter Findings

"No researcher would admit . . . bias . . . and some may not even be aware that they harbor . . . prejudices and have allowed their research to be vitiated thereby."[13] However, of interest to any integrative reviewer must be how much the experimenter's frame of reference influenced his or her findings. Three measures of potential bias were used: 1. Whether the researcher appeared in his or her literature review to favor the practice of ability grouping; 2. Whether the researcher mentioned any potential negative consequences of ability grouping on within school segregation by race or socioeconomic background; and 3. Whether the researcher was personally involved in Gifted and Talented Education.

After reading the introduction, experimental rationale and particularly the literature review of many studies, the experimenter's opinion about the efficacy of ability grouping frequently became very clear. The effect size findings for studies where that position was apparent are reported in Table 35. Those findings are astounding. Overall, those who favor ability grouping find a positive effect for ability grouping, while those who appear to oppose ability grouping find the opposite, = 0.08 (n = 349) and = 0.24 (n = 346) respectively. Also of note is the findings that the proponents and opponents of ability grouping are equally represented in the research findings.

When only achievement results were investigated, the findings of the two groups of researchers were even more disparate. As shown in Table 36, those who appear to favor ability grouping found that grouping students by ability improved their achievement by +0.25 standard deviations while those who oppose ability grouping found that it hurt achievement by -0.40--a difference of six-tenths of a standard deviation. In other words, the average ability grouped student was either at the sixtieth percentile compared to the average student in the heterogeneous classroom, or he was at the thirty-fourth percentile compared to the student in the heterogeneous class.

As shown in Table 37, researchers who favored ability grouping and those who opposed ability grouping both found negative self-concept implications for the practice, but those who opposed the practice found more than a three times greater negative effect, -0.06 (n = 162) and -0.21 (n = 231) respectively.

Another measure of an experimenter's position on ability grouping was obtained by recording which studies mentioned any potential negative consequences of ability grouping on racial or socioeconomic integration within a school. As shown in Table 38, 83 percent (n = 584) of the effect sizes in this meta-analysis came from studies which did not cite any potentially negative consequences of ability grouping such as creating within-the-school resegregation of students according to race, ethnicity or socioeconomic status. Overall, for researchers who did cite such potential consequences, the findings were almost three times more negative than the findings of those researchers who did not cite such implications, -0.17 (n = 123) and -0.06 (n = 584) respectively. As shown in Table 39, for cognitive effect sizes, the difference was wide, and the findings in the opposite direction: those researchers who mentioned potentially negative consequences of ability grouping found that ability grouping harmed achievement, while those who did not mention such effects found achievement enhanced by ability grouping, -0.23 (n = 30) and +0.05 (n = 241) respectively. As shown in Table 40, the findings of those who mentioned negative effects of ability grouping and those who did not mention such effects were the same on affective measures, -0.15 (n = 89 and n = 313 respectively).

Still another measure of the experimenter's position vis-a-vis ability grouping might be the experimenter's personal involvement in Gifted and Talented Education. The difference in effect size findings for those studies in which the experimenter was a "gifted and talented" educator are shown in Tables 41 and 42. On this measure of potential bias, as well as

the previous two measures, the differences were striking. The difference in findings by educators of the Gifted and Talented was positive and very large, while for educators who did not cite any such involvement the average findings were negative, +0.50 (n = 45) and -0.07 (n = 160) respectively. On affective measures those not involved in Gifted and Talented Education found negative effects twice as large as those so involved, -0.20 (n = 255) and -0.11 (n = 104) respectively.

Study Quality Findings

A common criticism leveled against the technique of meta-analysis is that it it indiscriminate; that is, that good studies and poor studies are lumped together and the effect sizes from excellent studies are lost in a sea of findings from studies which fail to meet the standards of excellence demanded by traditional reviewers of experimental literature. In this analysis, a measure of internal validity was used for overall quality. In order to be rated high on interval validity, a study had to use random assignment of pupils to the experimental and control groups. Some studies which used random assignment had other problems and their ratings dropped to average. A study which failed to randomize but did use other methods for determining beginning equivalence of the experimental and control groups and which exhibited no other serious internal validity threats was rated average. Studies rated as low exhibited a wide variety of methodological flaws--almost the entire gamut mentioned in any beginning research methodology text. Fourteen percent of the effect sizes in this analysis came from studies rated as high on internal validity, 54 percent from studies

rated as average, and 32 percent for studies rated as exhibiting low internal validity.

Table 43 shows the differences found in high and low internal validity studies investigating the achievement of students who were ability grouped versus those who were not so grouped. On cognitive measures, for high internal validity studies the students who were ability grouped scored almost two-tenths of a standard deviation lower than the ungrouped students, while for studies exhibiting low internal validity the findings were almost exactly the opposite, -0.19 (n = 44) and +0.18 (n = 116). As shown in Table 44, for affective measures both high and low internal validity studies found an effect size less than one-tenth of a standard deviation in favor of the students who were not ability grouped, -0.05 (n = 53) and -0.06 (n = 95) respectively.

Table 45 summarizes the relationships among this set of researcher and research quality variables. The Spearman correlation coefficients shown are non-parametric, and thus do not depend on the data assuming a normal distribution or interval scales. This correlation, rs, gives an idea of the degree of the relationship between pairs of variables, but only an idea. The reader is reminded that these data are heavily interdependent. It can be noted, however, that researchers who appear to favor ability grouping are likely (r = .62) to be involved in Gifted and Talented education and unlikely (r = -.19) to mention any potential negative in-school effects of ability grouping.

Summary Findings

The major findings of this study were that when students who were ability grouped were compared to similar students who were not ability grouped.

1. The overall outcome scores of the ability grouped students were lower.

2. The overall cognitive outcome scores for ability grouped and non-ability grouped students did not differ.

3. The overall affective outcome scores of the ability grouped students were lower.

4. The Content Area Skills outcome scores of the ability grouped students were higher.

5. The Attitude Toward Subject Matter scores of the ability grouped students were higher.

6. The Academic Self-Concept scores of the ability grouped students were lower.

7. The Self-Esteem scores of the ability grouped students were lower.

8. There was no overall pattern of grade level effects.

9. The affective outcome scores of the females were twice as negative as the males' scores.

10. For high and low ability students, the cognitive outcome scores of the ability grouped students were higher.

11. For average ability students, the cognitive outcome scores of the ability grouped students were lower.

12. For all ability levels, the affective outcome scores of the ability grouped students were lower.

13. There were subject matter differences in outcomes:

--The Reading scores, both cognitive and affective, of the ability grouped students were higher.

--The Mathematics scores, both cognitive and affective, of the ability grouped students were lower.

--The Language Arts scores, both cognitive and affective, of the ability grouped students were lower.

--The Social Studies scores, both cognitive and affective, of the ability grouped students were lower.

14. For students who spent a small percentage of the school day in ability groups, the scores, both cognitive and affective, of ability grouped students were lower.

15. For students who were ability grouped for two years or more, the scores, both cognitive and affective, of the ability grouped students were higher.

An examination of the characteristics of studies and their relationship to study findings revealed the following:

1. The number of studies on ability grouping has declined in recent years.

2. The average effect size reported has become more negative in recent years.

3. Larger effect sizes were found in published studies than in unpublished studies.

4. Students in average-size schools scored lower when ability grouped than students in either large or small schools.

5. Students in schools in lower socioeconomic communities who were ability grouped scored higher than students in higher socioeconomic communities who were ability grouped.

6. Students in schools in heterogeneous socioeconomic school communities who were ability grouped scored lower.

An examination of possible measures of experimenter bias and of study quality and their relationships to study findings revealed the following:

A. Researchers who appeared to favor ability grouping found that ability grouping increased achievement and other cognitive measures while researchers who appeared to oppose ability grouping found heterogeneous grouping more beneficial to students.

B. Both researchers who appeared to favor ability grouping and those who appeared to oppose ability grouping found that ability grouped students scored lower on self-concept, self-esteem and on affective measures.

C. Researchers who cited potential negative implications for ability grouping on in-school integration found greater negative overall outcomes than researchers who did not cite such negative consequences.

D. Researchers who were involved in Gifted and Talented Education found that ability grouping increased student outcomes while researchers who were not so involved found that ability grouped students scored lower.

E. Studies exhibiting high levels of internal validity found negative effects for ability grouping, while studies with low levels of internal validity found positive effects for ability grouping.

Conclusions

The interpretive discussions in this section are presented to make clearer both the meaning of these data and the implications of these data for school practice.

This study re-analyzed the recent experimental literature on ability grouping for the purpose of answering the question "Does ability grouping work?" The broad answer to that general question, now based on more than a quick glance at the previous research, is "No." This conclusion should have the power to move school policy makers in the direction of more effective and more just schools.

The strengths of this study are the general strengths of the meta-analytic method: it is quantitative and reports results quantitatively. It is based on all the research reported on ability grouping between 1967 and 1983. It is not based on "before the fact" judgments about the relationship of study quality to study findings. It is based on a very large data set, and it presents the findings of this large data set in a way that should be easily understood by those who might use its results.

The weaknesses of this study are, for the most part, the general weaknesses of the meta-analytic technique, and two criticisms should be considered in relation to this study: that meta-analysis is a case of Garbage-In-Garbage-Out, and that the data used in a meta-analysis are heavily inter-dependent, thus rendering inferential statistical methods inoperable.

The criticism that meta-analysis is a case of "Garbage-In-Garbage-Out"; that is, that a meta-analysis uses low standards of judgment about

the studies integrated is best answered after the analysis, not before. In the case of the ability grouping research, study quality was indeed related to study outcomes. The largest average negative effect size was found in studies of only average quality. High quality studies also found overall negative effects for ability grouping. Only low quality studies found overall positive results for ability grouping students for instruction. If only high and average quality studies had been used, the average overall effects of ability grouping would have been found to be even more negative than is reported earlier in this study. The overall effect size excluding the low quality studies would have been -0.16 (n = 489) rather than -0.08 (n = 720).

It should be remembered, however, that although some of the studies analyzed in this study were of poor design, these studies were undertaken in real schools with real pupils, and internal validity was limited by the demands of real world research. Very few schools with ability grouping programs are willing to allow a researcher to assign students randomly to programs which the school personnel believe will not benefit students. Parents are also unwilling to permit their children to be denied the "benefits" of ability grouping. The most consistent finding in the ability grouping research over the past fifty years is that parents and teachers believe that ability grouping works and is good for students.

The criticism that meta-analyses are conducted on large data sets in which multiple results are derived from the same study rendering the data non-independent is a very cogent one. This study dealt with that problem by reporting few correlations among variables, and by reporting the results of inferential statistics with cautions to the reader. Methods

are being developed to deal with the problem of interdependence, but no fully satisfactory solution has yet been reported. Readers of this study must decide how much reliance to place on inferential statistics and how much of the understanding of these data must be arrived at by inspection of the data and consideration of their implications for school practice.

The results of this study indicate that ability grouping, while favored by most teachers and entrenched in the public schools of the United States, does not (except in some very specific circumstances) improve student achievement and has potentially serious negative self-concept consequences.

What, then, should the school person, interested in solving the pedagogical problems of individual differences while at the same time responding to the evidence of the lack of overall efficacy of ability grouping, do?

1. Consider carefully the evidence on ability grouping.

2. Inform parents that, contrary to common myths, their students, even their high ability students, are not always well served when placed in ability groups.

3. Teach all elementary students in heterogeneous classrooms providing for teaching some subjects in ability groups, others interest groups, others in deliberately formed heterogeneous or working toward group mastery of the subject matter.

4. Make certain that, in the judgment of the teacher and the students, a wide variety of abilities and talents are actually valued.

5. When ability grouping does occur, follow guidelines which will help ensure that equality of educational opportunity is truly available to all our students.

In summation, the experimental evidence indicates that ability grouping does not improve overall student achievement and does damage overall student self-concept. We ought to be seeking policies and programs which enhance educational outcomes and which promote fairness in educational processes. Ability grouping does neither.

TABLES

Table 1

Experimental Variables

Variable	Question Asked About Each Study
Grade Level	In what grade level was the study conducted?
Sex	What percentage of females were in the experimental and control groups?
Race/Ethnicity	What percentage of minority students were in the experimental and control groups?
	What was the race or ethnicity of those minority students?
Ability Level	In what ability level were the students in the experimental and control groups?
Subject	What subject matter was taught in the experimental and control classes?
Time	What percentage of the school day was spent in the ability-grouped class?
	How many weeks were spent in the ability-grouped class?

Table 2

Study Characteristics

Characteristic	Question Asked About Each Study
Publication Information	In what year was the study reported?
	In what publication form were the data reported?
	In what size school was the study conducted?
	What was the socio-economic status of the school community where the study was conducted?
	In what type of community was the school where the study was conducted?
Curriculum Information	What pattern of curriculum organization was employed at the school where the study was conducted?
	Were curriculum changes made for the high group?
	Were curriculum changes made for the average group?
	Were curriculum changes made for the low group?
	Into how many ability levels were the students divided?
Measurement Information	What type of measurment instrument was used in the study?
	How reactive was the measurement instrument?[a]
	What was the size of the experimental information?
	How were the students assigned to the experimental group?

[a]See discussion of reactivity on p. 186.

Table 2 (Continued)

Characteristic	Question Asked About Each Study
Experimenter Information	Did the experimenter appear to favor ability grouping?
	Did the experimenter mention any negative socio-economic or racial consequences of ability grouping?
	Did the experimenter say he or she was involved in gifted and talented education?
Overall Study Information	How high was the internal validity of the study? i.e., what was the rated quality of the study?

Table 3

Number of Effect Sizes Per Study

Mean number of effect sizes	15
Maximum number of effect sizes	63
Minimum number of effect sizes	1
Standard deviation	13
Total number of effect sizes for entire analysis	720

Table 4

Average Effect Size Per Study

Mean Δ at the study level	0
Minimum Δ at the study level	-1.107
Maximum Δ at the study level	+1.301
Std. error	.005
Std. dev.	.122
n of Δ	50

Table 5

Overall Effect Size on All Outcome Measures

Mean:	-.08
Minimum:	-2.31
Maximum:	2.57
Std. dev.	.642
n of s	720
95% Confidence interval	-.128 to -.034

Table 6

Cognitive Measures Effect Size

Mean Δ	.01
Minimum	-2.31
Maximum	2.57
Standard deviation	.70
n of Δs	272
95% Confidence Interval	-.071 to .098

Table 7

Affective Measures Effect Size

Mean Δ	-.15
Minimum	-1.57
Maximum	1.99
Standard deviation	.60
n of Δs	402
95% Confidence Interval	-.206 to -.089

Table 8

Effect Sizes for Various Outcome Types

Outcome Type	$\bar{\Delta}$	Range Minimum	Range Maximum	95% Confidence Interval	s	n of Δs
Cognitive Outcomes						
Content Area Knowledge	-.04	-2.06	1.71	-.18 to .10	.7	98
Content Area Skills	.24	-0.50	1.33	.06 to .47	.5	18
Content Area Achievement	.05	-1.79	2.57	-.14 to .23	.7	66
Reading	.000	-2.31	1.50	-.18 to .18	.8	75
Grade Point Average	-.19	-.89	.28	-.41 to .03	.4	13
Other Cognitive Outcomes	1.37	1.08	1.67	-2.38 to 5.12	.4	2
Affective Outcomes						
Attitude in General	.12	-.85	.45	-.56 to .80	.6	5
Attitude Toward Subject Matter	.22	-.21	1.03	.05 to .40	.3	15
Attitude Toward School	.03	-.28	.82	-.19 to .25	.4	13
Attitude Toward Others	.09	-1.42	1.76	-.11 to .29	.6	41
Attitude Toward Ability Grouping	.07	-.82	.74	-.12 to .27	.5	28
Global Self-Concept	-.04	-.90	1.13	-.15 to .06	.4	56
Academic Self-Concept	-.30	-1.47	1.49	-.41 to .20	.6	133
Sense of Efficacy	-.30	-.32	-.29	-.494 to .11	.02	2
Self-Esteem	-.36	-1.6	1.99	-.535 to -.189	.7	67
Motivation	-.52	-.52	-.52	-.52 to -.52	.7	1
Aspirations	.01	-.18	.29	-.52 to -.52	.2	18
Other Affective Measures	-.01	-1.42	1.70	-.08 to .10	.7	18

Table 8 (Continued)

Outcome Type	$\bar{\Delta}$	Range Minimum	Range Maximum	95% Confidence Interval	s	n of Δs
Other Outcome Measures						
Sociometric Status	-.07	-1.73	.89	-.24 to .09	.5	39
Creativity	1.22	1.22	1.22	1.22 to 1.22	-	1
College Entrance	-.16	-.59	.44	-.56 to .23	.4	6

[a]s = the standard deviation of $\bar{\Delta}$.

Table 9

Overall Effect Sizes by Grade Level of Students

Grade Level of Students	$\bar{\Delta}$	n of Δs
K-3	-.04	133
4-6	-.11	196
Middle School	.08	24
Junior High	-.08	298
High School	-.15	69

Table 10

Cognitive Effect Sizes by Grade Level of Students

Grade Level	$\bar{\Delta}$	n of Δs
K-3	.004	84
4-6	-.20	58
Middle School	.08	24
Junior High	.22	87
High School	-.34	19

Table 11

Affective Effect Sizes by Grade Level of Students

Grade Level	$\bar{\Delta}$	n of Δs
K-3	-.11	49
4-6	-.11	120
Junior High	-.20	191
High School	-.05	42

Table 12

Cognitive Effect Sizes by Sex of Student

	Percentage of Females	$\bar{\Delta}$	n of Δs
PCT $Females_{AG}$	0	-.12	194
PCT $Females_{HG}$	0		
PCT $Females_{AG}$	100	-.31	2
PCT $Females_{HG}$	100		

Note: PCT $Females_{AG}$ = Percentage of females in the ability group scores; PCT $Females_{HG}$ = percentage of females in the heterogeneous group.

Table 13

Affective Measures Effect Sizes by Sex

	Percentage	$\bar{\Delta}$	n of Δs
PCT $Females_{AG}$	0	-.22	166
PCT $Females_{HG}$	0		
PCT $Females_{AG}$	100	-.11	69
PCT $Females_{HG}$	100		

Note: PCT $Females_{AG}$ = percentage of females in the ability group scores; PCT $Females_{HG}$ = percentage of females in the heterogeneous group scores.

Table 14

Cognitive Effect Sizes by Percentage of Minority Students

	Percentage Minority Students	$\bar{\Delta}$	n of Δs
$PCTMIN_{AG}$ $PCTMIN_{HG}$	0-5 0-5	-0.02	195
$PCTMIN_{AG}$ $PCTMIN_{HG}$	95-100 95-100	.31	18

Note: $PCTMIN_{AG}$ = percentage of minority students' scores in the ability grouped scores; $PCTMIN_{HG}$ = percentage of minority students' scores in the heterogeneously grouped scores.

Table 15

Affective Effect Sizes by Percentage of Minority Students

	Percentage Minority Students	$\bar{\Delta}$	n of Δs
$PCTMIN_{AG}$ $PCTMIN_{HG}$	0-5 0-5	-.17	291
$PCTMIN_{AG}$ $PCTMIN_{HG}$	95-100 95-100	+.24	15

Note: $PCTMIN_{AG}$ = percentage of minority students scores in the ability grouped scores; $PCTMIN_{HG}$ = percentage of minority student scores in the heterogeneous scores.

Table 16

Cognitive Effect Sizes by Ability Level of Students

Ability Level of Students	$\bar{\Delta}$	n of Δs
High	.16	82
Average	-.45	36
Low	.18	72

Table 17

Affective Effect Sizes by Ability Level of Students

Ability Level of Students	$\bar{\Delta}$	n of Δs
High	-.09	208
Average	-.15	33
Low	-.35	106

Table 18

Cognitive Effect Sizes for Various Subject Matter

Subject	$\bar{\Delta}$	n of Δs
Reading	.06	83
Math	-.11	63
Language Arts	-.01	33
Social Studies	-.06	40
Science	1.71	4
Other	.003	17

Table 19

Affective Effect Sizes for Various Subject Matter

Subject	$\bar{\Delta}$	n of Δs
Reading	.16	6
Math	-.45	22
Language Arts	-.12	20
Social Studies	-.08	67
Other	-.25	12

Table 20

Cognitive Effect Sizes by Grade Level, Subject Matter and Ability Level of Students

Grade	Subject	Ability Level of Students	$\bar{\Delta}$	n of Δs
K-3	Reading	a	.01	54
		High to High	-1.11	16
		Low to Low	.68	28
		All to All	-.06	10
K-3	Math		.02	13
		High to High	-.22	1
		Low to Low	.62	7
		All to All	-.77	5
K-3	Language Arts		-.12	14
		All to All	-.12	14
K-3	Social Studies		.42	2
		Low to Low	.42	2
K-3	Other		.15	1
		All to All	.15	1
4-6	Reading		-.005	25
		High to High	.01	4
		Average to Average	-.38	4
		Low to Low	-.06	4
		All to All	.12	13
4-6	Math		-.57	21
		Average to Average	-.73	16
		All to All	-.07	5

[a]The line marked a = the average effect size for all the levels below it.

Note: High = the average effect size for high ability students; Average = the average effect size for average ability students; Low = the average effect size for low ability students; All = the average effect size for studies which compared all levels of students.

(Table 20 Continued)

Grade	Subject	Ability Level of Students	$\bar{\Delta}$	n of Δs
K-4	Language Arts		.03	12
		High to High	.84	4
		Average to Average	-.20	4
		Low to Low	-.55	4
Middle School	Math		1.24	2
		Low to Low	1.24	2
Middle School	All		-.02	22
		All to All	-.02	22
Junior High	Reading		1.02	4
		High to High	1.02	4
Junior High	Math		.08	27
		High to High	.14	16
		Low to Low	-.32	4
		All to All	.15	7
Junior High	Language Arts		.16	7
		High to High	.16	7
Junior High	Social Studies		-.05	33
		High to High	.10	15
		Average to Average	-.18	9
		Low to Low	-.18	9
Junior High	Science		1.71	4
		High to High	1.71	4
Junior High	All		1.37	4
		High to High	1.37	4
Junior High	Other		.20	8
		High to High	1.37	4
		Low to Low	-.98	4

(Table 20 Continued)

Grade	Subject	Ability Level of Students	$\bar{\Delta}$	n of Δs
High School	Social Studies		-.33	
		Low to Low	.14	4
		All to All	0	1
High School	All Subject		-.53	6
		High to High	-.10	1
		Average to Average	-.52	1
		Low to Low	-.76	2
		All to All	-.53	2
High School	Other		-.21	8
		High to High	.62	2
		Average to Average	-.19	2
		Low to Low	-1.13	2
		All to All	-.14	2

Table 21

Affective Effect Sizes by Grade Level, Subject Matter and Ability Level of Students

Grade Level	Subject	Ability Level of Students[a]	$\bar{\Delta}$	n of Δs
K-3	Math	a	0	1
		All to All	0	1
K-3	All		-.12	48
		High to High	-.12	20
		Average to Average	-.02	12
		Low to Low	-.18	16
4-6	Reading		.16	6
		All to All	.16	6
4-6	All		-.12	109
		High to High	-.14	82
		Average to Average	-.35	4
		Low to Low	-.04	21
		All to All	.42	2
4-6	Other		-.27	5
		High to High	-.21	4
		All to All	0	1
Junior High	Math		-.48	21
		High to High	-.61	10
		Low to Low	-.37	10
		All to All	-.14	1
Junior High	Language Arts		-.12	20
		High to High	.55	10
		Low to Low	.79	10

[a]The line marked a = the average effect size for all the levels below it.

Note: High = the average effect size for high ability students; Average - the average effect size for average ability students; Low = the average effect size for low ability students; All = the average effect size for studies which compared all levels of students.

Table 21 (Continued)

Grade Level	Subject	Ability Level of Students	$\bar{\Delta}$	n of Δs
Junior High	Social Studies		-.17	43
		High to High	.24	16
		Average to Average	-.19	9
		Low to Low	-.52	18
Junior High	All		-.17	107
		High to High	-.10	49
		Low to Low	-.75	21
		All to All	.05	37
High School	Social Studies		.08	24
		High to High	-.06	8
		Average to Average	-.20	8
		Low to Low	.48	8
High School	All		-.22	11
		High to High	-.21	9
		Low to Low	-.28	2
High School	All		-.24	7
		All to All	-.24	7

Table 22

Cognitive Effect Sizes by Percentage of School Day in Ability Group

Percentage of Day in Ability Group	$\bar{\Delta}$	n of Δs
17	-.27	53
50	.16	1
84	-.53	6
100	.22	154

Table 23

Affective Effect Sizes by Percentage of School Day in Ability Group

Percent of Day in Ability Group	$\bar{\Delta}$	n of Δs
10	-.70	2
17	-.33	12
50	-.24	13
60	.16	4
66	-.39	40
84	.08	24
100	-.13	256

Table 24

Overall Effect Sizes by Year of Study Publication

Year	$\bar{\Delta}$	% of Δs	n of Δs
1967	-.04	10	74
1968	.22	13	93
1969	-.02	17	123
1970	.38	< 1	3
1971	.05	8	57
1972	-.28	24	169
1973	-.19	1	8
1974	.51	1	7
1975	-.22	7	49
1976	-.30	4	28
1977	-.06	1	7
1978	-.14	5	33
1979	.27	1	4
1980	-.10	7	53
1981	-	-	-
1982	-	-	-
1983	-.25	2	12

Table 25

Overall Effect Size by Form of Publication

Form	$\bar{\Delta}$	n of Δs
Journal	-.25	108
Thesis	-.12	375
Unpublished	.06	237

Table 26

Cognitive Effect Sizes by Socioeconomic (SES) Status of the School Community

School SES	$\overline{\Delta}$	n of Δs
Upper and Upper Middle	.01	43
Middle	.13	11
Lower	.29	70
Heterogeneous	-.20	40

Table 27

Affective Effect Sizes by Socioeconomic Status of the School Community

School SES	$\overline{\Delta}$	n of Δs
Upper and Upper Middle	-.13	51
Middle	-.11	148
Lower	-	-
Heterogeneous	-.27	147

Table 28

Overall Effect Sizes by Geographic Locations of School

Location	$\bar{\Delta}$	% of Effect Sizes	n of Δs
Urban school	-.07	16	112
Inner city school	.26	9	61
Suburban school	-.13	40	273
Rural school	.06	10	66
Small town school	-.15	7	49
More than 1 type of location	-.20	18	121

Table 29

Overall Effect Sizes by Type of Measurement Instrument Used by Original Researcher

Type of Measurement Instrument	$\bar{\Delta}$	%	n of Δs
National published test	-.11	69	499
Experimenter developed test	.005	25	177
Small distribution test	-.17	< 1	3
School records	-.25	3	25
Other	.07	2	16

Table 30

Overall Effect Sizes by Reactivity of Measurement Instrument

Reactivity of Instrument	$\bar{\Delta}$	%	n of Δs
High reactivity	-.13	47	340
Somewhat reactive	-.09	15	104
Low reactivity	-.02	38	276

Table 31

Cognitive Effect Sizes by Reactivity Measurement Instrument

Reactivity of Instrument	$\bar{\Delta}$	n of Δs
High reactivity	.04	30
Low reactivity	.01	242

Table 32

Affective Effect Size by Reactivity Measurement Instrument

Reactivity of Instrument	$\bar{\Delta}$	n of Δs
High reactivity	-.16	291
Somewhat reactive	-.11	87
Low reactivity	-.13	24

Table 33

Cognitive Effect Sizes by Experimental Placement Method

Placement Method	$\overline{\Delta}$	n of Δs
Random	-.20	70
Matched pairs	.54	57
Covariance adjusted of intact groups	.16	1
Intact groups	-.09	132

Table 34

Affective Measures Effect Sizes by Experimental Placement Method

Placement Method	$\overline{\Delta}$	n of Δs
Random	-.04	48
Matched pairs	-.10	15
Covariance adjusted of intact groups	-.19	7
Intact groups	-.18	325

Table 35

Overall Effect Sizes by Experimenter's Apparent Bias

Position	$\bar{\Delta}$	n of Δs
Experimenter appears to favor ability grouping	.08	349
Experimenter appears to oppose ability grouping	-.24	346

Table 36

Cognitive Effect Sizes by Experimenter's Apparent Bias

Position	$\bar{\Delta}$	n of Δs
Experimenter appears to favor ability grouping	.25	170
Experimenter appears to oppose ability grouping	-.39	98

Table 37

Affective Effect Sizes by Experimenters' Apparent Bias

Position	$\bar{\Delta}$	n of Δs
Experimenter appears to favor ability grouping	-.06	162
Experimenter appears to oppose ability grouping	-.21	231

Table 38

Overall Effect Sizes by Researcher Citing Potential Negative Implications of Ability Grouping on In-School Integration

Experimenter Cites Potential Negatives	$\bar{\Delta}$	% of Δs	n of Δs
Yes	-.17	17	123
No	-.06	83	584

Table 39

Cognitive Effect Sizes by Researcher Citing Potential Negative Effects of Ability Grouping on In-School Integration

Researcher Cites Negatives	$\bar{\Delta}$	n of Δs
Yes	-.23	30
No	.05	241

Table 40

Affective Effect Sizes by Researcher Citing Potential Negative Effects of Ability Grouping on In-School Integration

Researcher Cites Negatives	$\bar{\Delta}$	n of Δs
Yes	-.15	89
No	-.15	313

Table 41

Cognitive Effect Sizes by Researcher's Involvement in "Gifted and Talented" Education (GAT)

Researcher Involved in GAT Education	$\overline{\Delta}$	n of Δs
Yes	.50	45
No	-.07	160

Table 42

Affective Effect Sizes by Researcher's Involvement in "Gifted and Talented" Education (GAT)

Researcher Involved in GAT Education	$\overline{\Delta}$	n of Δs
Yes	-.11	104
No	-.20	255

Table 43

Cognitive Effect Sizes by Internal Validity of Study

Quality	$\bar{\Delta}$	n of Δs
High	-.19	44
Average	-.08	112
Low	.18	116

Table 44

Affective Effect Sizes by Internal Validity of Study

Quality	$\bar{\Delta}$	n of Δs
High	-.05	53
Average	-.20	254
Low	-.06	95

Table 45

Spearman Correlations for Overall Effect Size with Various Measures of Potential Researcher Bias and Internal Validity of the Experimental Studies

	FAV	NEG	GAT	Internal Validity	Effect Size
Researcher appears to favor ability grouping (FAV)[b]	-	-.19[a]	.62[a]	-.23[a]	-.23[a]
Researcher cites potential negative implications of ability grouping on in-school segregation (NEG)		-	-.21[a]	-.05[a]	-.04
Researcher involved in Gifted and Talented Education (GAT)[b]			-	-.17[a]	-.10[a]
Internal validity of study[c]				-	.18[a]

[a] sig. = .001.

[b] These three variables were coded Yes = 1, No = 2.

[c] This variable was coded 1 = High, 2 = Average, 3 = Low.

ENDNOTES

[1]Theresa Koontz Noland, "The Effects of Ability Grouping: A Meta-Analysis of Research Findings," Ph.D. dissertation, University of Colorado, 1985.

[2]Sarane Spence Boocock, Sociology of Education: An Introduction (Boston: Houghton Mifflin Company, 1980), p. 143.

[3]Caroline Hodges Persell, Education and Inequality: A Theoretical and Empirical Synthesis (New York: The Free Press, 1977), p. 85.

[4]W. S. Miller and J. H. Otto, "Analysis of Experimental Studies in Homogeneous Grouping," Journal of Educational Research, 21 (1930), 95.

[5]Stephen C. Hey, "Some Consequences of Ability Grouping and Tracking in Schools," in People in Schools, ed. Blaine E. Mercer and Stephen C. Hey (Cambridge, Massachusetts, 1981), p. 228.

[6]Kenneth A. Feldman, "Using the Work of Others: Some Observations on Reviewing and Integrating," Sociology of Education, 44 (Winter 1971), 87.

[7]Herbert J. Walberg and Edward H. Hoertl, "Research Integration: Introduction and Overview," Evaluation in Education, 4 (1980), 8.

[8]Gene V. Glass, Barry McGaw, and Mary Lee Smith, Meta-Analysis in Social Research (Beverley Hills: Sage Publications, 1981), p. 102.

[9]Ibid., pp. 198-99.

[10]Walter R. Borg and Meredith D. Gall, Educational Research: An Introduction (New York: Longman, 1983), p. 19.

[11]Glass, McGaw, and Smith, p. 227.

[12]Ibid., p. 84.

[13]Kenneth Carlson, "Ways in Which Research Methodology Distorts Policy Issues," Urban Review, 11, No. 1 (1979), 5.

BIBLIOGRAPHY

Boocock, Sarane Spence. Sociology of Education: An Introduction. Boston: Houghton Mifflin Company, 1980.

Borg, Walter R., and Meredith D. Gall. Educational Research: An Introduction. New York: Longman, 1983.

Carlson, Kenneth. "Ways in Which Research Methodology Distorts Policy Issues." Urban Review, 11, No. 1 (1979), 3-14.

Feldman, Kenneth A. "Using the Work of Others: Some Observations on Reviewing and Integrating." Sociology of Education, 44 (Winter 1971), 86-102.

Glass, Gene V., Barry McGaw, and Mary Lee Smith. Meta-Analysis in Social Research. Beverley Hills: Sage Publications, 1981.

Mercer, Blaine E., and Stephen C. Hey, eds. People in Schools. Cambridge, Massachusetts: Schenkman Publishing Company, 1981.

Noland, Theresa Koontz. "The Effects of Ability Grouping: A Meta-Analysis of Research Findings." Ph.D. dissertation, University of Colorado, 1985.

Persell, Caroline Hodges. Education and Inequality: A Theoretical and Empirical Synthesis. New York: The Free Press, 1977.

Walberg, Herbert J., and Edward H. Haetel. "Research Integration: An Introduction and Overview." Evaluation in Education, 4 (1980), 5-10.